

AD\_\_\_\_\_

Award Number: W81XWH-12-1-0013

TITLE: Whole Exome Analysis of Early Onset Alzheimer's Disease

PRINCIPAL INVESTIGATOR: Margaret Pericak-Vance

CONTRACTING ORGANIZATION: University of Miami  
Miami, FL 33136

REPORT DATE: April 2015

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188		
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
<b>1. REPORT DATE</b> April 2015		<b>2. REPORT TYPE</b> Annual		<b>3. DATES COVERED</b> 1Mar2014 - 31Mar2015	
<b>4. TITLE AND SUBTITLE</b> Whole Exome Analysis of Early Onset Alzheimer's Disease			<b>5a. CONTRACT NUMBER</b> W81XWH-12-1-0013		
			<b>5b. GRANT NUMBER</b>		
			<b>5c. PROGRAM ELEMENT NUMBER</b>		
<b>6. AUTHOR(S)</b> Margaret A. Pericak-Vance, Ph.D.  E-Mail: mpericak@med.miami.edu			<b>5d. PROJECT NUMBER</b>		
			<b>5e. TASK NUMBER</b>		
			<b>5f. WORK UNIT NUMBER</b>		
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> University of Miami, 1400 NW 10 <sup>th</sup> Avenue, Miami, FL 33136			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>		
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>		
			<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>		
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b> Approved for Public Release; Distribution Unlimited					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b> The primary focus toward identification of Alzheimer disease (AD) risk genes over the past five years has been testing the common disease common variant (CDCV) hypothesis through the use of genome-wide association studies (GWAS) in late onset Alzheimer disease (LOAD). While common variation clearly plays a role in AD there is a growing realization that the CDCV hypothesis is unlikely to explain all the genetic effect underlying AD. One alternative hypothesis invokes multiple rare variants (RV) in one or more genes, each with stronger individual effects than CDCV genes. We designed this project to test the rare variant hypothesis in AD by examining those cases with the most severe phenotype as determine by early onset (EOAD, cases with AAO < 60 years). Although there are three known EOAD genes (PS1, PS2 and APP) they account for only ~60-70% of familial EOAD and even less of sporadic EOAD. Thus, the majority of the genetics of EOAD remains unknown. Until now, large extended families with AD in multiple generations were necessary to identify variants of significant effect contributing to AD risk, however, with the advent of new genomic technologies such as high-throughput sequencing technology, small family aggregates and isolated cases, particularly those with an extreme phenotype of the disorder (such as early onset) can be used. Thus, we will utilize whole exome high-throughput sequencing to identify high risk AD variants that we will further characterize with respect to AD. We will examine both Caucasian and Caribbean Hispanic AD populations. Our two pronged approach includes structural characterization at the DNA level (Dr. Pericak-Vance), and analysis of Caribbean Hispanics (Dr. Richard Mayeux). Comparing across populations will be extremely useful. Specifically, high priority RVs identified through the whole exome analysis will be further explored with multiple strategies. We will also genotype the interesting variants in a large sample of late-onset (LOAD) cases to examine their involvement in all AD. We will thus prepare a list of high priority candidates for additional follow-up and functional analysis.					
<b>15. SUBJECT TERMS</b>  Nothing Listed					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>  UU	<b>18. NUMBER OF PAGES</b>  23	<b>19a. NAME OF RESPONSIBLE PERSON</b> USAMRMC
<b>a. REPORT</b> U	<b>b. ABSTRACT</b> U	<b>c. THIS PAGE</b> U			<b>19b. TELEPHONE NUMBER (include area code)</b>

## Table of Contents

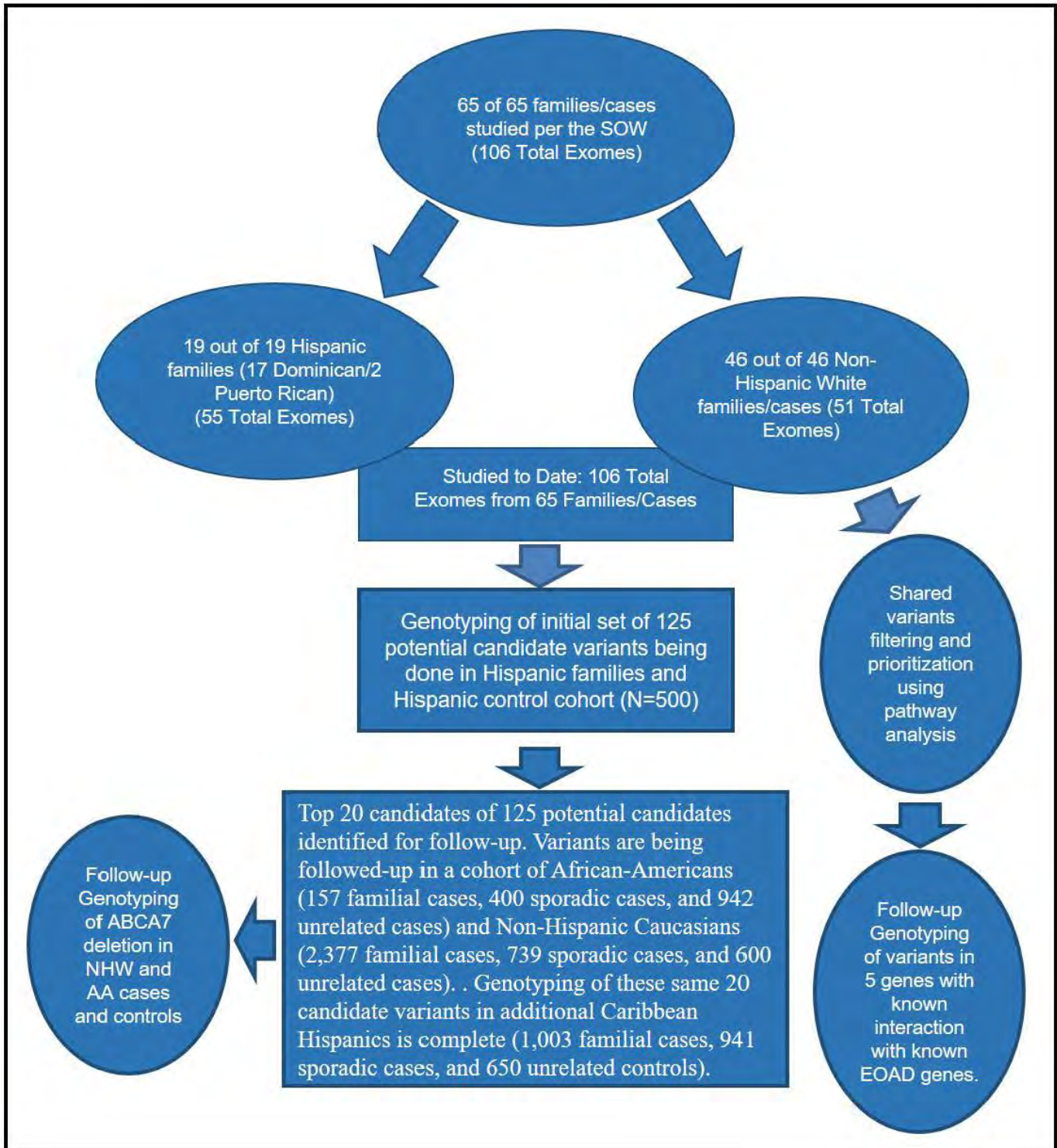
	<u>Page</u>
Introduction.....	4
Body.....	5
Key Research Accomplishments.....	7
Reportable Outcomes.....	8
Conclusion.....	9
Appendices.....	10

## **INTRODUCTION:**

The primary focus toward identification of Alzheimer disease (AD) risk genes over the past five years has been testing the common disease common variant (CDCV) hypothesis through the use of genome-wide association studies (GWAS) in late onset Alzheimer disease (LOAD). While common variation clearly plays a role in AD there is a growing realization that the CDCV hypothesis is unlikely to explain all the genetic effect underlying AD. One alternative hypothesis invokes multiple rare variants (RV) in one or more genes, each with stronger individual effects than CDCV genes. We designed this project to test the rare variant hypothesis in AD by examining those cases with the most severe phenotype as determined by early onset (EOAD, cases with AAO < 60 years). Although there are three known EOAD genes (PS1, PS2 and APP) they account for only ~60-70% of familial EOAD and even less of sporadic EOAD. Thus, the majority of the genetics of EOAD remains unknown. Until now, large extended families with AD in multiple generations were necessary to identify variants of significant effect contributing to AD risk, however, with the advent of new genomic technologies such as high-throughput sequencing technology, small family aggregates and isolated cases, particularly those with an extreme phenotype of the disorder (such as early onset) can be used. Thus, we will utilize whole exome high-throughput sequencing to identify high risk AD variants that we will further characterize with respect to AD. We will examine both Caucasian and Caribbean Hispanic AD populations. Our two pronged approach includes structural characterization at the DNA level (Dr. Pericak-Vance), and analysis of Caribbean Hispanics (Dr. Richard Mayeux). Comparing across populations will be extremely useful. Specifically, high priority RVs identified through the whole exome analysis will be further explored with multiple strategies. We will also genotype the interesting variants in a large sample of late-onset (LOAD) cases to examine their involvement in all AD. We will thus prepare a list of high priority candidates for additional follow-up and functional analysis.

**BODY:**

An overview of the analyses can be seen in Figure 1. Briefly 19 Hispanic families (N = 55 total exomes) and 46 non-Hispanic White (NHW) families and sporadic cases (N=51 total exomes) were filtered prioritizing rare, damaging variants. 125 variants found segregating in the Hispanic families were follow-up genotyped in a other family members and a large set of Hispanic cases and controls. 20 variants were then prioritized for further follow-up based on MAF and segregation, including a deletion in ABCA7 that was genotyped in in a large set of African-American (AA) and NHW cases and controls. A separate analysis that looked for shared rare, damaging variants across the NHW cases only was also performed and identified several other candidates for follow-up genotyping. This genotyping is part of the no-cost extension. Details of these analyses follow.



**Figure 1. Analysis Flowchart for project.**

### *WES and variant prioritization*

Whole exome sequencing (WES), quality control and variant calling, variant annotation, and variant filtering is complete on 55 samples submitted by Columbia University to the University of Miami. Additionally, WES and analysis of 51 samples from 46 multiplex families from The University of Miami and Vanderbilt University is complete. Identity-by-descent analysis of Hispanic families was also performed. Following these analyses, comparison of the candidate variants/genes shared across Hispanic families and NH-white cases was done. From these analyses, a list of 125 unique variants was prioritized for follow-up genotyping.

A brief overview of how each family was filtered individually and how variants for typing were prioritized follows:

- 1) Quality Filter per individual WES sample: VQSLOD>0, PL Score>100, Read Depth>6
- 2) Annotation of remaining variants with ANNOVAR
- 3) Remove variants with MAF>0.001 in EVS\_6500si and 1000G2012mar\_all and MAF>0.01 in HIHG internal controls
- 4) Keep variants with Autosomal dominant and X-linked dominant segregation in family
- 4) Exclude variant if not missense, Splicing, Stopgain, Stoploss, Nonframeshift Indel, or Frameshift Indel in refSeq gene annotation, Ensemble gene annotation, or UCSC Known gene annotation
- 5) Filter on deleteriousness based on a) damaging score in any of these 7 programs programs: Sift, Polyphen2\_HDIV, LRT, MutationTaster, MutationAssessor, or FATHMM and b) conservation based on a conserved score in any of these 3 programs: GERP, SiPhy or PhyloP
- 7) Apply IBD sharing results and require 100% sharing in Hispanic families with enough GWASed individuals
- 8) Genotype any variant passing above filters and in a known EOAD or LOAD
- 9) Interrogate shared variants and variants in shared genes across Hispanic Families and between Hispanic and NH-White Families by screening them for existence and potentially too high a MAF in dbSNP, EVS, 1000G updates, specific 1000G populations (EA, AA, AMR and ASN, and any population in UCSC), and cg69 (69 complete genomics exomes). Because of the large amount of candidate genes generated from filtering of the NH-White cases, a variant from the comparison of Hispanic and NH-White candidates was only carried forward for genotyping if the variant/gene passed this screening and was in 2+ Hispanic families and 2+ NH-White cases. Additionally, variants/genes still in 2+ Hispanic families after the screening were carried forward for genotyping.
- 8) Additional variants were selected by applying a 'secondary filter' to the Hispanic families in order to reduce single variant per family candidates:
  - remove any single nucleotide variant (SNV) with an rs# in dbSNP129-dbSNP137
  - remove all indels
  - remove families with greater than 50 variants remaining (families 1,171,386 and 419)
  - keep only variants predicted to be damaging in 3 or more of the 7 prediction programs used
  - NOTE: Candidate variants for the four removed families were selected based on shared variants/genes with other families.

### *Follow-up Genotyping of Top Candidate Variants*

261 Hispanic familial subjects from 19 pedigrees (145 affecteds and 116 unaffecteds) and 500 Hispanic non-familial subjects (382 healthy controls and 118 sporadic EOAD cases) were genotyped for these 125 top candidate variants. 101 of the variants passed all QC filters (13 variants failed genotyping and 11 were monomorphic in the dataset). For analysis of results of this follow-up genotyping we: 1) estimated familial and population frequencies of the variants in our follow-up cohort and 2) tested single SNV association with AD with 2 models using generalized estimation equations (GEE):

M1) AD~SNV+AGE+SEX

M2) AD~SNV+AGE+SEX+APOE

20 top candidate variants were identified from this follow-up genotyping, including a 44 base-pair deletion in ABCA7 that was further genotyped and Sanger sequenced separately. The other 19 variants include 8 SNVs that show perfect segregation with AD status in the families and are absent in population controls (Table 1). These variants are in the genes *MYO3A*, *AAAS*, *DICER1*, *YIPF1*, *ACAP1*, *LLGL2*, *BPIFB2*, and *ABCG2*. An additional 11 variants were identified as follow-up candidates based on them showing near complete segregation (absent in one or a few familial cases) and being absent in all familial and sporadic controls. These variants are in the genes *GPR26*, *ERCC6*, *OR5M9*, *DNAH3*, *MYOCD*, *KIF17*, *TICRR*, *PLXNB2*, *LAMA2*,

*SNRNP48*, and *GLB1L2*. These top 19 variants were then genotyped in large cohorts of Hispanics (1621 cases and 884 controls) (Table 2), African-Americans (157 familial cases, 400 sporadic cases, and 942 unrelated cases) and Non-Hispanic Caucasians (2,377 familial cases, 739 sporadic cases, and 600 unrelated cases). Genotyping in AA identified 1 case with the *DNAH3* variant; 3 controls with the *TICRR* variant (Age-of-Exams of 83, 62, and 67), 1 control with the *SNRNP48* variant (Age-of-Exam of 70), and 1 control with the *LAMA2* SNV (Age-of-Exam of 70). Genotyping in NHW cases and controls identified 1 familial case with the *KIF17* variant (Age-of-onset of 75) and another case in *PLXNB2* (Age-of-onset of 51). NOTE: Assays for *MYOCD*, *ACAP1*, *LLGL2*, *DICER1*, and *GPR26* could not be designed for follow-up genotyping in AA and NHW. Options for genotyping for these variants are under consideration.

**Table 1. 19 candidate variants from stage 1 validation genotyping of 125 candidate variants.**

CHR	POS	GENE	+ Family Cases	+ Family Ctrls	+ Sporadic Cases	+ Sporadic Controls	- Family Cases	- Family Cases with Early Onset	ESP FRQ*
10	26243836	<i>MYO3A</i>	4	0	0	0	0	0	--
12	53714383	<i>AAAS</i>	5	0	0	0	0	0	--
14	95574334	<i>DICER1</i>	5	0	0	0	0	0	--
1	54354580	<i>YIPF1</i>	5	0	0	0	0	0	--
17	7249740	<i>ACAP1</i>	5	0	0	0	0	0	9.3E-05
17	73552150	<i>LLGL2</i>	7	0	0	0	0	0	--
20	31606097	<i>BPIFB2</i>	5	0	0	0	0	0	--
4	89042878	<i>ABCG2</i>	5	0	0	0	0	0	--
10	1.25E+08	<i>GPR26</i>	3	0	0	0	1	0	--
10	50667132	<i>ERCC6</i>	4	0	0	0	1	0	--
11	56230523	<i>OR5M9</i>	6	0	1	0	1	0	--
16	21071613	<i>DNAH3</i>	2	0	1	0	1	0	--
17	12626229	<i>MYOCD</i>	5	0	0	0	1	0	--
1	21031194	<i>KIF17</i>	5	0	0	0	2	0	--
15	90145195	<i>TICRR</i>	5	0	0	0	2	0	1.0E-04
22	50727490	<i>PLXNB2</i>	5	0	0	0	2	0	--
6	1.3E+08	<i>LAMA2</i>	5	0	2	0	2	0	2.0E-04
6	7605638	<i>SNRNP48</i>	5	0	0	0	2	0	--
11	1.34E+08	<i>GLB1L2</i>	4	0	0	0	3	0	--

\* None of the SNVs were present in 1000 Genomes or dbSNP databases

**Table 2. Results of follow-up genotyping of top 19 Hispanic EOAD variants in 1621 Hispanic Cases and 884 Hispanic Controls**

Chr:Position	Gene	1621 Hispanic Cases MAF	884 Hispanic Controls MAF
10:26243836	<i>MYO3A</i>	0.0009482	0
12:53714383	<i>AAAS</i>	0	0
14:95574334	<i>DICER1</i>	0	0
1:54354580	<i>YIPF1</i>	0	0
17:7249740	<i>ACAP1</i>	0.0006324	0.001705
17:73552150	<i>LLGL2</i>	0.003519	0.005143
20:31606097	<i>BPIFB2</i>	0	0
4:89042878	<i>ABCG2</i>	0	0
10:125426036	<i>GPR26</i>	0	0
10:50667132	<i>ERCC6</i>	0.001261	0.002286
11:56230523	<i>OR5M9</i>	0.002527	0.001708
16:21071613	<i>DNAH3</i>	0	0
17:12626229	<i>MYOCD</i>	0	0
1:21031194	<i>KIF17</i>	0.0006333	0.001139
15:90145195	<i>TICRR</i>	0.0006321	0.001139
22:50727490	<i>PLXNB2</i>	0	0
6:129714360	<i>LAMA2</i>	0.001262	0.0005675
6:7605638	<i>SNRNP48</i>	0	0
11:134241025	<i>GLB1L2</i>	0	0

Follow-up genotyping of the 44 base pair deletion in *ABCA7* (rs142076058) found segregating in Hispanic Family 380 and the Reitz et al. 2013 (1) *ABCA7* risk SNP (rs115550680) was completed in cohorts of NHW and AA AD cases and controls. Results show the deletion is very rare in Non-Hispanic White cases and controls (0.12%). Testing in AA cases and controls, adjusting for age, sex, and APOE status, found the deletion to be significantly associated with disease ( $p=0.0002$ ,  $OR=2.13$  [95% CI:1.42-3.20]). The association was replicated in an independent dataset ( $p=0.0117$ ,  $OR=1.65$  [95% CI:1.12-2.44]). Joint analysis resulted in an effect size (OR) estimate = 1.81 ([95% CI:1.38-2.37]  $p=1.414 \times 10^{-5}$ ). The deletion is common in both AA cases (15.2%) and AA controls (9.74%). Linkage disequilibrium between the deletion and the Reitz et al. 2013 risk SNP is high at  $r^2 = 0.921$  and  $D' = 0.975$ .

#### *Non-Hispanic White Cases Only Analyses*

To search for rare variants contributing to risk for EOAD we performed Whole-Exome Sequencing (WES) in 50 Caucasian EOAD cases screened negative for *APP*, *PSEN1*, and *PSEN2*. Variant filtering for rare ( $MAF < 0.1\%$  in ExAC database) functional (nonsynonymous or loss-of-function(LOF)), damaging variants was performed. Damage prediction was performed using a Combined Annotation Dependent Depletion (CADD) score (2), with scores  $\geq 10$  considered damaging. Rare, damaging variants shared by multiple cases (+2) were then selected for follow-up protein-protein interaction analysis with known or suspected EOAD genes (*APP*, *PSEN1*, *PSEN2*, *GRN*, *MAPT*, *TREM2*, *SORL1*) using the program STRINGdb (3). This analysis identified 5 genes with the same rare, potentially damaging nonsynonymous or LOF variant in two or more EOAD cases and evidence for protein-protein interaction with a known EOAD gene (Table 3 and Figure 1). Several other cases have a rare nonsynonymous or LOF, potentially damaging variant in another variant in these 5 genes.

The 5 genes implicated are: *HSPG2* (interacts with *GRN* and *APP*), *CLSTN1* (interacts with *PSEN1* and *APP*), *DOCK3* (interacts with *PSEN1* and *PSEN2*), *PARK2* (interacts with *MAPT*, *PSEN1*, and *APP*), and *OGT* (interacts with *MAPT*) (Figure 1). Six cases have a variant in *HSPG2*, a gene in a LOAD susceptibility region and potentially involved in amyloidogenesis and tau aggregation in AD (4-6). Three cases have a variant in *DOCK3*, a gene shown to regulate amyloid- $\beta$  secretion, and associated with neurofibrillary tangles in AD brains (7,8). Two cases have shared variants in *CLSTN1*. Disruption of calyntenin-1-associated axonal transport of APP by mutations in *CLSTN1*, a known APP interactor (9,10), have been identified as a potential pathogenic mechanism of Alzheimer's (11). Moreover, *CLSTN1*'s potential as a regulator of synapse formation and neuronal development suggests other mechanisms through which it could be involved in development of dementia (12). Interestingly, *CLSTN1* interacts with another newly identified candidate gene from this analysis, *OGT*, which was found to have 12 EOAD cases carrying two separate frameshift insertions at the same position on Chromosome X (X: 70767666). Numerous studies exist linking *OGT* to neurodegeneration, including a study supporting its therapeutic potential due to its ability to prevent protein aggregation including reduction of formation of tau oligomers [13], and a study showing increased biochemical levels of *OGT* lead to slower cognitive decline and amyloid plaque formation in mice [14]. Finally, though variants in *PARK2* are the most frequent cause of autosomal recessive early-onset Parkinson's disease and juvenile Parkinson disease, Parkin has been shown to promote intracellular Abeta1-42 clearance [15], is upregulated in AD brains, and colocalizes with classic senile plaques and amyloid-laden vessels in AD brains [16], hinting at its potential involvement in Alzheimer's as well. Validation of these results using Sanger sequencing is underway.



**Table 3. Rare (MAF<0.001), nonsynonymous or loss-of-function variants (LOF) found in two or more EOAD cases. Additional rare, missense or LOF variants in these genes are also listed.**

Chr:Position	Ref/Alt Allele	rsID	Gene	Consequence	N Affected	MAF (ExAC)	CADD	STRINGdb Interaction
1:9795622	C/T	.	CLSTN1	missense	2	8.95E-05	15.92	APP, PSEN1, OGT
1:22211078	T/G	.	HSPG2	missense	2	N	16.49	APP, GRN
1:22163463	G/A	rs200225298	HSPG2	missense	1	2.47E-04	11.89	APP, GRN
1:22181841	C/T	.	HSPG2	missense	1	5.70E-05	22.9	APP, GRN
1:22183791	G/A	.	HSPG2	missense	1	8.13E-05	26.4	APP, GRN
1:22205523	C/T	rs150650673	HSPG2	missense	1	1.14E-04	28.6	APP, GRN
3:51251601	G/A	rs199600118	DOCK3	missense	2	5.80E-04	29.5	PSEN1, PSEN2
3:51312575	G/C	.	DOCK3	missense	1	N	29	PSEN1, PSEN2
6:161969982	G/+C	.	PARK2	frameshift	1	8.13E-06	35	APP, MAPT, PSEN1
6:161990390	C/G	rs72480423	PARK2	missense	3	1.95E-04	22.4	APP, MAPT, PSEN1
X:70767666	T/+C	.	OGT	frameshift	8	1.72E-04	14.86	MAPT, CLSTN1
X:70767666	T/+CC	.	OGT	frameshift	4	1.72E-04	14.86	MAPT, CLSTN1

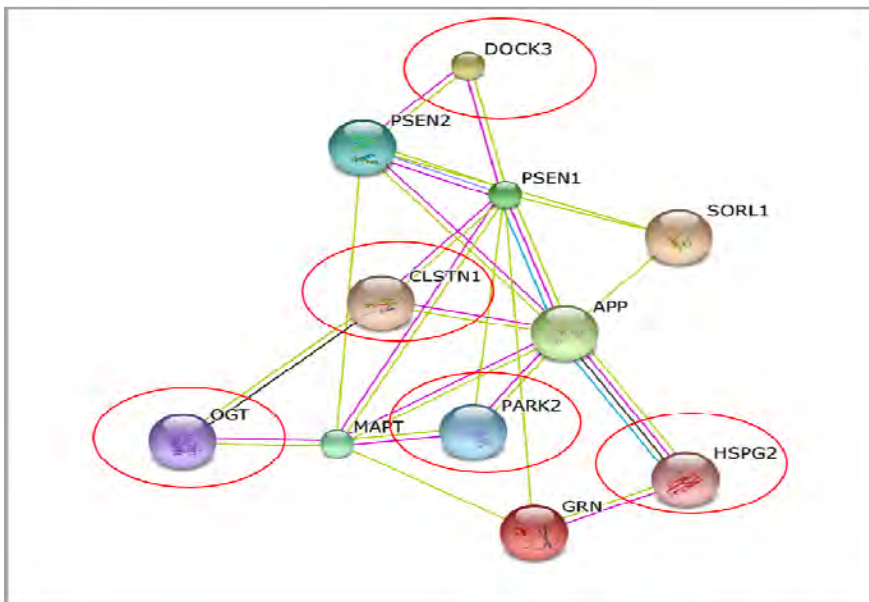


Figure 2. STRINGdb network analysis of top Non-white Hispanic EOAD candidate genes using known or suspected EOAD genes (APP, PSEN1, PSEN2, TREM2, MAPT, TREM2, SORL1, and GRN) as seed nodes. ‘Strings’ between genes represent evidence of protein-protein interaction between linked genes.

We additionally found several rare coding variants in known or suspected EOAD genes (Table 3), and are investigating a potential link to Parkinsonism and SORL1 (see Cuccaro ASHG 2015 abstract below).

Table 3. Variants in non-Hispanic White Cases in known EOAD genes.

N Cases	Age	Chr	Start	Ref	Alt	Gene
2	61,61	17	44101427	C	T	MAPT
1	52	17	44101427	C	T	MAPT
1	54	14	73637653	C	T	PSEN1
1	56	14	73637653	C	T	PSEN1
1	50	14	73664774	C	G	PSEN1
1	59	11	121414334	C	T	SORL1
1	55	11	121498300	C	T	SORL1
1	48	1	227071475	C	T	PSEN2
1	48	1	227075813	A	G	PSEN2

## REFERENCES

- [1] Reitz et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E  $\epsilon$ 4, and the risk of late-onset Alzheimer disease in African Americans. *JAMA*. 2013 Apr 10;309(14):1483-92.
- [2] Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–5. doi:10.1038/ng.2892.
- [3] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 2013;41:D808–15. doi:10.1093/nar/gks1094.
- [4] Kehoe P, Wavrant-De Vrieze F, Crook R, Wu WS, Holmans P, Fenton I, et al. A full genome scan for late onset Alzheimer’s disease. *Hum Mol Genet* 1999;8:237–45.
- [5] Hiltunen M, Mannermaa a., Thompson D, Easton D, Pirskanen M, Helisalmi S, et al. Genome-wide linkage disequilibrium mapping of late-onset Alzheimer’s disease in Finland. *Neurology* 2001;57:1663–8. doi:10.1212/WNL.57.9.1663.
- [6] Castillo G, Ngo C, Cumming J, Wight TN, Snow AD. Perlecan Binds to the  $\beta$ -Amyloid Proteins ( $A\beta$ ) of Alzheimer’s Disease, Accelerates  $A\beta$  Fibril Formation, and Maintains  $A\beta$  Fibril Stability. *J Neurochem* 1997;2452–65.
- [7] Goedert M, Jakes R, Spillantini MG, Hasegawa M, Smith MJ, Crowther RA. Assembly of microtubule-associated protein tau into Alzheimer-like filaments induced by sulphated glycosaminoglycans. *Nature* 1996;383:550–3. doi:10.1038/383550a0.
- [8] Chen Q, Kimura H, Schubert D. A novel mechanism for the regulation of amyloid precursor protein metabolism. *J Cell Biol* 2002;158:79–89. doi:10.1083/jcb.200110151.
- [9] Chen Q, Yoshida H, Schubert D, Maher P, Mallory M, Masliah E. Presenilin binding protein is associated with neurofibrillary alterations in Alzheimer’s disease and stimulates tau phosphorylation. *Am J Pathol* 2001;159:1597–602. doi:10.1016/S0002-9440(10)63005-2.
- [10] Araki Y, Tomita S, Yamaguchi H, Miyagi N, Sumioka A, Kirino Y, et al. Novel cadherin-related membrane proteins, Alcadeins, enhance the X11-like protein-mediated stabilization of amyloid beta-protein precursor metabolism. *J Biol Chem* 2003;278:49448–58. doi:10.1074/jbc.M306024200.
- [11] Vagnoni A, Perkinton MS, Gray EH, Francis PT, Noble W, Miller CCJ. Calsyntenin-1 mediates axonal transport of the amyloid precursor protein and regulates  $a\beta$  production. *Hum Mol Genet* 2012;21:2845–54. doi:10.1093/hmg/dds109.
- [12] Um JW, Pramanik G, Ko JS, Song MY, Lee D, Kim H, et al. Calsyntenins Function as Synaptogenic Adhesion Molecules in Concert with Neurexins. *Cell Rep* 2014;6:1096–109. doi:10.1016/j.celrep.2014.02.010.
- [13] Yuzwa SA, Shan X, Macauley MS, Clark T, Skorobogatko Y, Vosseller K, Vocadlo DJ. Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat Chem Biol*. 2012;8:393-9. doi: 10.1038/nchembio.797.
- [14] Yuzwa SA, Shan X, Jones BA, Zhao G, Woodward ML, Li X, Zhu Y, McEachern EJ, Silverman MA, Watson NV, Gong CX, Vocadlo DJ1. Pharmacological inhibition of O-GlcNAcase (OGA) prevents cognitive decline and amyloid plaque formation in bigenic tau/APP mutant mice. *Mol Neurodegener*. 2014;9:42. doi: 10.1186/1750-1326-9-42.
- [15] Burns MP, Zhang L, Rebeck GW, Querfurth HW, Moussa CE. Parkin promotes intracellular Abeta1–42 clearance. *Hum Mol Genet* 2009;18:3206–3216.
- [16] Witte ME, Bol JG, Gerritsen WH, van der Valk P, Drukarch B, van Horssen J, Wilhelmus MM. Parkinson’s disease-associated parkin colocalizes with Alzheimer’s disease and multiple sclerosis brain lesions. *Neurobiol Dis*. 2009 Dec;36:445-52. doi: 10.1016/j.nbd.2009.08.009.

## KEY RESEARCH ACCOMPLISHMENTS:

- Variant calling and quality control processing of these samples completed on 55 Hispanic individuals submitted by Columbia and 51 NH-White samples from the University of Miami and Vanderbilt University.
- Analysis (variant annotation and filtering) completed on samples of 55 Hispanic individuals submitted by Columbia and 51 NH-White samples from the University of Miami and Vanderbilt University.
- Identity-by-descent analysis of Hispanic families is complete.
- Identification of 125 top candidate variants for follow-up genotyping is complete.
- Genotyping of 125 top candidate variants in the Hispanic families and a cohort of 500 Hispanic cases controls is complete.
- Analysis of the 125 top candidate variants in the Hispanic families and a cohort of 500 Hispanic cases and controls is complete, with 20 top candidates identified for follow-up, including a 44 base-pair deletion in the known LOAD gene *ABCA7*.
- Follow-up of these 20 top candidates in large cohorts of Hispanic, AA and NHW AD cases and controls helped prioritize the top candidate genes and confirmed the association of the deletion in *ABCA7* to increased risk of AD.
- Identification of rare coding variants in *SORL1*, *PSEN1*, and *MAPT* in EOAD cases including a potential link between *SORL1* and Parkinsonism.
- Identification of 5 candidate early-onset Alzheimer disease genes (*HSPG2*, *DOCK3*, *OGT*, *CLSTN1*, and *PARK2*) through identification of NHW EOAD cases with shared rare coding variants with damaging potential in genes interacting with known EOAD genes.

## REPORTABLE OUTCOMES:

Platform Presentation (Appendix I): Kunkle BW, Kohli MA, Vardarajan BN, Reitz C, Naj AC, Whitehead PL, Martin ER, Beecham GW, Gilbert JR, Farrer LA, Haines JL, Schellenberg GD, Mayeux RP, Pericak-Vance MA, Alzheimer's Disease Genetics Consortium. Whole-exome sequencing in early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer-related genes and processes. The 63rd Annual Meeting of the American Society of Human Genetics (ASHG), Boston, MA, October 22-26, 2013.

Platform Presentation (Appendix II): Reitz C, Kunkle BW, Vardarajan BN, Kohli MA, Naj AC, Whitehead PL, Perry WR, Martin ER, Beecham GW, Gilbert JR, Farrer LA, Haines JL, Schellenberg GD, Pericak-Vance MA, Mayeux RP, Alzheimer's Disease Genetics Consortium. Whole-exome sequencing of Hispanic early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer-related genes. The American Academy of Neurology (AAN) 66<sup>th</sup> Annual Meeting, Philadelphia, PA, April 26-May 3, 2014.

Poster Presentation (Appendix III): Margaret Pericak-Vance, Christiane Reitz, Brian W. Kunkle, Badri N. Vardarajan, Martin A. Kohli, Adam C. Naj, Patrice L. Whitehead, William Perry, Eden Martin, Gary Beecham, John Gilbert, Lindsay A. Farrer, Jonathan Haines, Gerard D. Schellenberg, Richard Mayeux. Whole-exome sequencing of Hispanic early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer's-related genes. The Annual Alzheimer's Association International Conference, Copenhagen, 2014.

Poster Presentation at ADPD 2015 (Appendix VI): Margaret A. Pericak-Vance, Brian W. Kunkle, Badri Vardarajan, Patrice L. Whitehead, Sophie Rolati, Eden R. Martin, John R. Gilbert, Gary W. Beecham, Richard P. Mayeux, Jonathan L. Haines. Whole-exome sequencing in early-onset Alzheimer disease cases identifies several novel candidate genes. 12<sup>th</sup> Annual Conference on Alzheimer's Disease and Parkinson's Disease's Conference, Nice, FRA, March 18-22 2015.

Poster Presentation at ADPD 2015 (Appendix V): Carney RM, et al. Novel and known mutations in SORL1, PSEN1, and PSEN2 genes are found in multiplex Alzheimer's disease families with varying age of onset and pathological presentations. 12<sup>th</sup> Annual Conference on Alzheimer's Disease and Parkinson's Disease's Conference, Nice, FRA, March 18-22, 2015.

Platform Presentation at AAIC 2015 (Appendix VI): Gary W. Beecham, PhD; Brian W. Kunkle, PhD, MPH; Badri Vardarajan, PhD; Patrice L. Whitehead, BS; Sophie Rolati, MS; Eden R. Martin, PhD; John R. Gilbert, PhD. Whole-Exome Sequencing in Early-Onset Alzheimer Disease Cases Identifies Novel Candidate Genes. The Annual Alzheimer's Association International Conference, Washington, D.C. July 18-23, 2015.

Poster Presentation at AAIC 2015 (Appendix VII): Holly N. Cukier, PhD; Brian W. Kunkle, PhD, MPH; Sophie Rolati, MS; Kara L. Hamilton-Nelson, MPH; Martin A. Kohli, PhD; Beth A. Dombroski, PhD; Badri N. Vardarajan, PhD; Patrice L. Whitehead, BS; Derek J. Van Booven, BS; Eden R. Martin, PhD; Gary W. Beecham, PhD; Lindsay A. Farrer, PhD; Michael L. Cuccaro, PhD; Jeffery M. Vance, MD, PhD; Richard Mayeux, MD, MSc; John R. Gilbert, PhD; Regina M. Carney, MD; Goldie S. Byrd, PhD; Jonathan L. Haines, PhD; Gerald D. Schellenberg, PhD; Margaret A. Pericak-Vance, PhD; Rosalyn Lang, PhD and Alzheimer Disease Genetics Consortium. ABCA7 Deletion Associated with Alzheimer's Disease in African Americans. The Annual Alzheimer's Association International Conference, Washington, D.C. July 18-23, 2015.

Poster Presentation accepted at ASHG 2015 (Appendix VIII): Brian W. Kunkle, Badri Vardarajan, Patrice L. Whitehead, Sophie Rolati, Eden R. Martin, John R. Gilbert, Richard P. Mayeux, Jonathan L. Haines, Margaret A. Pericak-Vance, Gary W. Beecham. Whole-exome sequencing identifies novel candidate genes for early-onset Alzheimer disease. The 65<sup>th</sup> Annual Meeting of The American Society of Human Genetics, Baltimore, MD, October 6-10, 2015.

Poster Presentation accepted at ASHG 2015 (Appendix IX): Cuccaro ML, Carney RM, Kunkle BW, Vance JM, Whitehead PL, Gilbert JR, Vardarajan BN, Haines JL, Mayeux R, Pericak-Vance MA. SORL1 mutations and Parkinsonian features in early onset Alzheimer's disease families. The 65th Annual Meeting of The American Society of Human Genetics, Baltimore, MD, October 6-10, 2015.

## **Manuscripts**

ABCA7 deletion associated with Alzheimer's disease in African Americans. *Submitted to JAMA*

Whole-exome sequencing of Hispanic families identifies novel candidate genes for early-onset Alzheimer disease. *In preparation.*

Whole-exome sequencing identifies novel candidate genes for early-onset Alzheimer disease. *In preparation.*

SORL1 mutations and Parkinsonian features in early onset Alzheimer's disease families. *In preparation.*

## CONCLUSION:

Mutations in *APP*, *PSEN1* and *PSEN2* lead to familial EOAD and accounting for 60-70% of familial EOAD and ~11% of EOAD overall, leaving the majority of genetic risk for this form of Alzheimer disease unexplained. We performed Whole-Exome Sequencing (WES) on 55 individuals in 19 Caribbean Hispanic EOAD families and 51 Non-Hispanic White EOAD cases previously screened negative for *APP*, *PSEN1* and *PSEN2* to search for rare variants contributing to risk for EOAD. Variants were filtered for segregating, conserved and functional rare variants (MAF<0.1%) assuming both autosomal and X-linked dominant models. 125 rare, segregating, conserved and functional variants passed our stringent filtering criteria for selection of follow-up genotyping candidates. These variants have undergone follow-up genotyping for segregation in the families and for presence in a cohort of 500 Hispanic cases and controls. 20 top candidate variants were identified from this follow-up genotyping, including a 44 base-pair deletion in the known LOAD gene *ABCA7* that was associated with risk of AD in several follow-up cohorts. They include 8 variants that show perfect segregation with AD status in the families and are absent in population controls. These variants are in the genes *MYO3A*, *AAAS*, *DICER1*, *YIPF1*, *ACAP1*, *LLGL2*, *BPIFB2*, and *ABCG2*. An additional 11 variants were identified as follow-up candidates based on them showing near complete segregation (absent in one or a few familial cases) and being absent in all familial and sporadic controls. These variants are in the genes *GPR26*, *ERCC6*, *OR5M9*, *DNAH3*, *MYOCD*, *KIF17*, *TICRR*, *PLXNB2*, *LAMA2*, *SNRNP48*, and *GLB1L2*. Follow-up of these 20 top candidates in large cohorts of Hispanic, AA and NHW AD cases and controls helped prioritize several top candidate genes and confirmed the association of the deletion in *ABCA7* to increased risk of AD. We also identified several rare coding variants in *SORL1*, *PSEN1*, and *MAPT* in EOAD cases and are investigating a potential link between *SORL1* and Parkinsonism in *SORL1* carriers. Finally, we identified 5 additional candidate EOAD genes (*HSPG2*, *DOCK3*, *OGT*, *CLSTN1*, and *PARK2*) through identification of NHW EOAD cases with shared rare coding variants with damaging potential in genes interacting with known EOAD genes. Follow-up sequencing and genotyping of the variants in these 5 genes is ongoing.

## APPENDICES:

### Appendix I:

## **Whole-exome sequencing in early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer-related genes and processes**

Brian W. Kunkle<sup>1</sup>, Martin A. Kohli<sup>1</sup>, Badri N. Vardarajan<sup>2</sup>, Christiane Reitz<sup>2</sup>, Adam C. Naj<sup>3</sup>, Patrice L. Whitehead<sup>1</sup>, Eden R. Martin<sup>1</sup>, Gary W. Beecham<sup>1</sup>, John R. Gilbert<sup>1</sup>, Lindsay A. Farrer<sup>3</sup>, Jonathan L. Haines<sup>4</sup>, Gerard D. Schellenberg<sup>5</sup>, Richard P. Mayeux<sup>2</sup>, Margaret A. Pericak-Vance<sup>1</sup>, and The Alzheimer's Disease Genetics Consortium.

<sup>1</sup> John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA

<sup>2</sup> Taub Institute of Research on Alzheimer's Disease, Columbia University, New York, NY, USA

<sup>3</sup> School of Medicine, Boston University, Boston, MA, USA

<sup>4</sup> Center for Human Genetics Research, Vanderbilt University, Nashville, TN, USA

<sup>5</sup> Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

### **Background**

Mutations in *APP*, *PSEN1* and *PSEN2* lead to familial, early-onset Alzheimer disease (EOAD). These mutations account for only 60-70% of familial EOAD and ~11% of EOAD overall, leaving the majority of genetic risk for the most severe form of Alzheimer disease unexplained.

### **Methods**

We performed Whole-Exome Sequencing in Caribbean Hispanic and Caucasian EOAD families previously screened negative for *APP*, *PSEN1*, and *PSEN2* to search for rare variants contributing to risk for EOAD. 60 individuals in 21 families were sequenced using the Agilent 50Mb kit on an Illumina HiSeq2000. Variant filtering for segregating, conserved and functional rare variants (MAF<0.1%) was performed on the 21 families assuming both autosomal-dominant and X-linked dominant models. Filtered loci were examined for implication as AD candidate genes from GWAS or in biologically relevant KEGG Pathways. Variants were also followed up for association with AD in 13,748 individuals (7,652 affected) from the Alzheimer's Disease Genetics Consortium (ADGC) genotyped on the exome chip, which included 195,039 variants with MAF<2%. Enrichment analysis of the variant list was conducted using DAVID.

### **Results**

984 variants in 886 genes passed our stringent filtering criteria, including 63 genes with rare segregating, conserved and functional variants in two or more families. A frameshift mutation in *ABCA7* and a missense variant in *ZCWPW1* are present in one of the 23 GWAS-confirmed Alzheimer disease candidate genes. Seven variants are in AD KEGG Pathway genes (*BID*, *CYC1*, *ITPR1*, *ITPR2*, *LRP1*, *ATP2A1*), including two variants in *LRP1*, a gene involved in AD through its roles in cholesterol transport and  $\beta$ -amyloid modulation. Follow up in ADGC exome chip association results comparing EOAD vs. late-onset AD identified 13 of our filtered genes with suggestive associations ( $P<10^{-3}$ ), including *ITM2C* ( $P=1.22\times 10^{-4}$ ), a gene known to inhibit the processing of *APP* by blocking access to alpha- and beta-secretase. Enrichment analysis of the list of rare conserved, functional variants showed significant, Benjamini FDR-adjusted enrichment for several AD-related processes including the 'ECM-receptor interaction' and 'ABC transporters' KEGG pathways; GO terms including 'homophilic cell adhesion' and 'microtubule-based movement'; and multiple INTERPRO 'cadherin' classes.

### **Conclusion**

Exome sequencing of EOAD pedigrees identified multiple rare segregating variants with potential roles in AD pathogenesis, several of which were shared in two or more families.

## **Whole-exome sequencing of Hispanic early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer-related genes**

C. Reitz<sup>1</sup>, B. W. Kunkle<sup>2</sup>, B. N. Vardarajan<sup>1</sup>, M. A. Kohli<sup>2</sup>, A. C. Naj<sup>3</sup>, P. L. Whitehead<sup>2</sup>, W. R. Perry<sup>2</sup>, E. R. Martin<sup>2</sup>, G. W. Beecham<sup>2</sup>, J. R. Gilbert<sup>2</sup>, L. A. Farrer<sup>3</sup>, J. L. Haines<sup>4</sup>, G. D. Schellenberg<sup>5</sup>, M. A. Pericak-Vance<sup>2</sup>, R. P. Mayeux<sup>1</sup>, Alzheimer's Disease Genetics Consortium 1) Taub Institute for Research on Alzheimer's Disease, Columbia University, New York, NY, USA; 2) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA; 3) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 4) School of Medicine, Boston University, Boston, MA, USA; 5) Center for Human Genetics Research, Vanderbilt University, Nashville, TN, USA.

**OBJECTIVE:** To identify novel early-onset Alzheimer disease (EOAD) candidate genes.

**BACKGROUND:** Mutations in *APP*, *PSEN1* and *PSEN2* lead to familial EOAD and accounting for 60-70% of familial EOAD and ~11% of EOAD overall, leaving the majority of genetic risk for this form of Alzheimer disease unexplained.

**DESIGN/METHODS:** We performed Whole-Exome Sequencing (WES) on 55 individuals in 19 Caribbean Hispanic EOAD families previously screened negative for *APP*, *PSEN1* and *PSEN2* to search for rare variants contributing to risk for EOAD. Variants were filtered for segregating, conserved and functional rare variants (MAF<0.1%) assuming both autosomal and X-linked dominant models. Filtered loci were examined for implication as AD candidate genes by comparison to: late-onset Alzheimer (LOAD) susceptibility genes, biologically relevant Alzheimer KEGG Pathway genes, candidate genes from 45 WESed NH-White EOAD cases, and results of an Alzheimer's Disease Genetics Consortium (ADGC) exome chip association study.

**RESULTS:** 2,225 variants in 1,531 genes passed our stringent filtering criteria, including 308 genes with rare segregating, conserved and functional variants in two or more families. Frameshift insertions-deletions in *ABCA7* and *HLA-DRB1*, a nonframeshift deletion in *RIN3*, and missense variants in *DSG2* and *PICALM*, all LOAD susceptibility genes, were discovered. 11 AD KEGG Pathway genes have variants, including *LRP1*, a gene involved in cholesterol transport and  $\beta$ -amyloid modulation. 83 variant carrying genes are in 2+ Hispanic and 2+ Non-white Hispanic families, including the AD-relevant *HLA-A* (associated with earlier age-at-onset), *CHST15* (a potential modulator of A $\beta$  toxicity), and *NOTCH4* (a presenilin pathway gene). Exome chip results identified variants in *MICA* encoding the *HLA-A* gene and previously associated with LOAD in a small study, as having suggestive association ( $p=9.10 \times 10^{-4}$ ). One family has variants in both *HLA-A* and *MICA*.

**CONCLUSIONS:** Exome sequencing of Hispanic EOAD pedigrees identified multiple rare segregating variants with potential roles in AD pathogenesis, several of which were shared in two or more families.



## Appendix III:

### **Whole-exome sequencing of Hispanic early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer's-related genes.**

Margaret Pericak-Vanceemail, Christiane Reitz, Brian W. Kunkle, Badri N. Vardarajan, Martin A. Kohli, Adam C. Naj, Patrice L. Whitehead, William Perry, Eden Martin, Gary Beecham, John Gilbert, Lindsay A. Farrer, Jonathan Haines, Gerard D. Schellenberg, Richard Mayeux

**Background:** Mutations in APP, PSEN1 and PSEN2 lead to familial, early onset Alzheimer disease (EOAD) and account for only 60-70% of familial EOAD and ~11% of EOAD overall, leaving the majority of genetic risk for this severe form of Alzheimer disease unexplained.

**Methods:** To identify novel early-onset Alzheimer disease candidate genes we performed Whole-Exome Sequencing (WES) on 55 individuals in 19 Caribbean Hispanic EOAD families previously screened negative for APP, PSEN1, and PSEN2 to search for rare variants contributing to risk for EOAD. Variants were filtered for segregating, conserved and functional rare variants (MAF<0.1%) assuming both autosomal and X-linked dominant models. Filtered loci were examined for implication as AD candidate genes by comparison to: late-onset Alzheimer (LOAD) susceptibility genes, biologically relevant Alzheimer KEGG Pathway genes, candidate genes from 45 WESed NH-White EOAD cases, and results of an Alzheimer's Disease Genetics Consortium (ADGC) exome chip association study.

**Results:** 2,225 variants in 1,531 genes passed our stringent filtering criteria, including 308 genes with rare segregating, conserved and functional variants in two or more families. Frameshift insertions-deletions in ABCA7 and HLA-DRB1, a nonframeshift deletion in RIN3, and missense variants in DSG2 and PICALM, all LOAD susceptibility genes, were discovered. 11 AD KEGG Pathway genes have variants, including LRP1, a gene involved in AD through its roles in cholesterol transport and  $\beta$ -amyloid modulation. 83 variant carrying genes are in 2+ Hispanic and 2+ Non-white Hispanic families, including the AD-relevant HLA-A (associated with earlier AD age-at-onset), CHST15 (a potential modulator of Abeta toxicity), and NOTCH4 (a presenelin pathway gene). Exome chip results showed the variant carrying gene MICA, which encodes the HLA-A gene and has been associated with LOAD in a small study, as having suggestive association ( $p=9.10 \times 10^{-4}$ ). One family has variants in both HLA-A and MICA.

**Conclusions:** Exome sequencing of Hispanic EOAD pedigrees identified multiple rare segregating variants with potential roles in AD pathogenesis, several of which were shared in two or more families.

## **Whole-exome sequencing in early-onset Alzheimer disease cases identifies several novel candidate genes**

Margaret A. Pericak-Vance<sup>1</sup>, Brian W. Kunkle<sup>1</sup>, Badri Vardarajan<sup>2</sup>, Patrice L. Whitehead<sup>1</sup>, Sophie Rolati<sup>1</sup>, Eden R. Martin<sup>1</sup>, John R. Gilbert<sup>1</sup>, Gary W. Beecham<sup>1</sup>, Richard P. Mayeux<sup>2</sup>, Jonathan L. Haines<sup>3</sup>.

<sup>1</sup>John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA

<sup>2</sup>Taub Institute of Research on Alzheimer's Disease, Columbia University, New York, NY, USA

<sup>3</sup>Institute for Computational Biology, Case Western Reserve University, Cleveland, OH, USA

### **Objectives (35/250)**

Mutations in *APP*, *PSEN1* and *PSEN2* lead to early-onset Alzheimer disease (EOAD). These mutations account for ~11% of EOAD overall, leaving the majority of genetic risk for the most severe form of Alzheimer disease unexplained.

### **Methods (67/250)**

We performed Whole-Exome Sequencing (WES) in 50 Caucasian EOAD cases previously screened negative for *APP*, *PSEN1*, and *PSEN2* to search for rare variants contributing to risk for EOAD. Variant filtering for functional, damaging rare variants (MAF<0.1%) was performed. Genes with shared (2+ cases with the same variant), damaging variants were examined for interactions with known EOAD genes (*APP*, *PSEN1*, *PSEN2*, *SORL1*, *GRN*, *MAPT*) and *APOE* using STRINGdb.

### **Results (134/250)**

176 genes had rare functional variants shared in two or more cases. 46 of these genes were prioritized for their damaging potential, defined by their shared rare variants having a Combined Annotation Dependent Depletion (CADD) score in the top 10% of all variants. Gene network analysis of these 46 genes with known EOAD genes and *APOE* identified five top candidate genes: *HSPG2* (interacts with *GRN*, *APOE*, and *APP*), *CLSTN1* (interacts with *PSEN1* and *APP*), *DOCK3* (interacts with *PSEN1* and *PSEN2*), and the *APOE* interactors *SAR1B* and *STAT1*. 5 cases have a variant in *HSPG2*, a gene potentially involved in amyloidogenesis and tau aggregation in AD, while 4 cases have a variant in *DOCK3*, a gene expressed exclusively in the central nervous system and associated with neurofibrillary tangles in AD brains.

### **Conclusions (13/250)**

WES of EOAD cases identified several genes with potential roles in AD pathogenesis.

## Appendix V:

### **Novel and known mutations in SORL1, PSEN1, and PSEN2 genes are found in multiplex Alzheimer's disease families with varying age of onset and pathological presentations**

RM Carney et al.

#### **Objectives**

To identify causative AD-related mutations in 50 Caucasian multiplex families with at least one individual with early-onset Alzheimer's disease (EOAD).

#### **Methods**

Variants in the *SORL1*, *PSEN1*, and *PSEN2* genes were identified using an established whole-exome sequencing (WES) pipeline. Clinical characteristics of affected members were obtained via review of medical and research records.

**Results.** A novel *SORL1* mutation that is predicted to be damaging (T588I) was found in all 4 affected individuals in one family (age of onset [AOO] range 59-82; three with APOE genotype 3/4, one with 3/3). A second family carried a previously reported AD *SORL1* mutation (T749M), with 3 affected individuals with the mutation (AOO55-84, all APOE 3/3), and 1 affected individual without the mutation (AOO76, APOE 3/4).

A known *PSEN1* AD mutation (A79V) was identified in two families, with AOO54 years (APOE 3/4) and 56 years (APOE 3/4). A rare *PSEN1* variant (R269G) predicted to be damaging was identified in a third family (AOO50 years, APOE 3/3).

A single individual (AOO48, APOE 3/3) represented a compound heterozygote for 2 variants predicted to be damaging in *PSEN2* (R71W, M174V).

A large range of AOO was seen within the families, with some members classified as EOAD and other as LOAD. On autopsy, Lewy bodies were seen in an EOAD individual carrying *SORL1* T749M without clinical Parkinsonism.

#### **Conclusions**

Mutations were found in 6/50 families. The presence of an APOE-4 allele may account for disease status in one affected non-carrier (T749M). No prominent atypical clinical features were identified.

## Appendix VI:

### **Whole-Exome Sequencing in Early-Onset Alzheimer Disease Cases Identifies Novel Candidate Genes**

**Gary W. Beecham, PhD<sup>1</sup>**; Brian W. Kunkle, PhD, MPH<sup>1</sup>; Badri Vardarajan, PhD<sup>2</sup>; Patrice L. Whitehead, BS<sup>1</sup>; Sophie Rolati, MS<sup>1</sup>; Eden R. Martin, PhD<sup>1</sup>; John R. Gilbert, PhD<sup>1</sup>; Richard Mayeux, MD, MSc<sup>2</sup>; Jonathan L. Haines, PhD<sup>3</sup> and Margaret A. Pericak-Vance, PhD<sup>1</sup>, (1)University of Miami, Miami, FL, USA, (2)Columbia University, New York, NY, USA, (3)Case Western Reserve University, Cleveland, OH, USA

#### **Abstract**

**Background:** Mutations in *APP*, *PSEN1* and *PSEN2* lead to early-onset Alzheimer disease (EOAD). These mutations account for ~11% of EOAD overall, leaving the majority of genetic risk for the most severe form of Alzheimer disease unexplained.

**Methods:** We performed Whole-Exome Sequencing (WES) in 50 Caucasian EOAD cases previously screened negative for *APP*, *PSEN1*, and *PSEN2* to search for rare variants contributing to risk for EOAD. Variant filtering for functional, damaging rare variants (MAF<0.1%) was performed. Genes with shared (2+ cases with the same variant), damaging variants were examined for interactions with known EOAD genes (*APP*, *PSEN1*, *PSEN2*, *SORL1*, *GRN*, *MAPT*) and *APOE* using STRINGdb.

**Results:** 176 genes had rare functional variants shared in two or more cases. 46 of these genes were prioritized for their damaging potential, defined by their shared rare variants having a Combined Annotation Dependent Depletion (CADD) score in the top 10% of all variants. Gene network analysis of these 46 genes with known EOAD genes and *APOE* identified five top candidate genes: *HSPG2* (interacts with *GRN*, *APOE*, and *APP*), *CLSTN1* (interacts with *PSEN1* and *APP*), *DOCK3* (interacts with *PSEN1* and *PSEN2*), and the *APOE* interactors *SAR1B* and *STAT1*. 5 cases have a variant in *HSPG2*, a gene potentially involved in amyloidogenesis and tau aggregation in AD, while 4 cases have a variant in *DOCK3*, a gene expressed exclusively in the central nervous system and associated with neurofibrillary tangles in AD brains.

**Conclusions:** WES of EOAD cases identified several genes with potential roles in AD pathogenesis.

## Appendix VII:

### **ABCA7 Deletion Associated with Alzheimer's Disease in African Americans**

**Holly N. Cukier, PhD<sup>1</sup>**; Brian W. Kunkle, PhD, MPH<sup>1</sup>; Sophie Rolati, MS<sup>1</sup>; Kara L. Hamilton-Nelson, MPH<sup>1</sup>; Martin A. Kohli, PhD<sup>1</sup>; Beth A. Dombroski, PhD<sup>2</sup>; Badri N. Vardarajan, PhD<sup>3</sup>; Patrice L. Whitehead, BS<sup>1</sup>; Derek J. Van Booven, BS<sup>1</sup>; Eden R. Martin, PhD<sup>1</sup>; Gary W. Beecham, PhD<sup>1</sup>; Lindsay A. Farrer, PhD<sup>4</sup>; Michael L. Cuccaro, PhD<sup>1</sup>; Jeffery M. Vance, MD, PhD<sup>1</sup>; Richard Mayeux, MD, MSc<sup>3</sup>; John R. Gilbert, PhD<sup>1</sup>; Regina M. Carney, MD<sup>1</sup>; Goldie S. Byrd, PhD<sup>5</sup>; Jonathan L. Haines, PhD<sup>6</sup>; Gerald D. Schellenberg, PhD<sup>2</sup>; Margaret A. Pericak-Vance, PhD<sup>1</sup>; Rosalyn Lang, PhD<sup>5</sup> and Alzheimer Disease Genetics Consortium, (1)University of Miami, Miami, FL, USA, (2)University of Pennsylvania, Philadelphia, PA, USA, (3)Columbia University, New York, NY, USA, (4)Boston University, Boston, MA, USA, (5)North Carolina A&T State University, Greensboro, NC, USA, (6)Case Western Reserve University, Cleveland, OH, USA

#### **Abstract**

**Background:** The *ATP-binding cassette, sub-family A (ABC1), member 7 (ABCA7)* gene has been implicated as a risk factor in Alzheimer's disease (AD) in both African American and Caucasian populations. However, the effect in African Americans is significantly higher, comparable to that found in *APOE ε4*. Furthermore, the underlying damaging allele(s) conveying the strong genome-wide signal has yet to be revealed.

**Methods:** We performed custom next generation sequencing across the *ABCA7* region on 41 African American individuals with AD and 48 control African Americans carrying the previously reported risk allele, rs115550680. Using Agilent custom capture, probes were designed across a 150 kb genomic area that includes *ABCA7* as well as eight flanking genes and a small nuclear RNA. Samples were run on the Illumina HiSeq sequencing machine. Data processing was performed with Casava, GATK, and BWA, and deletions were identified with Pindel. Our top priority alterations were confirmed by Sanger sequencing and validated in two African American cohorts – HIHG (cases: 482, controls: 565) and ADGC (cases: 746, controls: 1,063). Analyses for each cohort were adjusted for gender, and age (controls >65).

**Results:** 1,120 SNVs were detected by sequencing. 11 variants had different frequencies in cases and controls ( $p < 0.1$ ). In addition, a 44 base pair deletion (rs142076058) was identified in all 89 cases and controls, signifying that it could be in high linkage disequilibrium with the risk allele. The deletion was significantly correlated with disease in the HIHG cohort ( $p = 0.020$ , odds ratio = 1.54, 95% confidence interval 1.07-2.22). This finding was then replicated in the ADGC cohort ( $p = 0.006$ , odds ratio = 1.50, 95% confidence interval 1.12-2.01). The deletion falls in the 14<sup>th</sup> exon of *ABCA7* and results in a frameshift and truncating mutation (p.Arg578Alafs) that could interfere with protein function. When we investigated this same deletion in our cohort of European ancestry (>3,000 samples), the deletion was only found in 5 individuals, 4 of whom also carried the risk allele, and did not correlate with disease.

**Conclusions:** This deletion in *ABCA7* could represent an ethnically-specific pathogenic alteration in Alzheimer's disease.

## **Whole-exome sequencing identifies novel candidate genes for early-onset Alzheimer disease**

Brian W. Kunkle, Badri Vardarajan<sup>2</sup>, Patrice L. Whitehead<sup>1</sup>, Sophie Rolati<sup>1</sup>, Eden R. Martin<sup>1</sup>, John R. Gilbert<sup>1</sup>, Richard P. Mayeux<sup>2</sup>, Jonathan L. Haines<sup>3</sup>, Margaret A. Pericak-Vance<sup>1</sup>, Gary W. Beecham<sup>1</sup>.

<sup>1</sup>John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA

<sup>2</sup>Taub Institute of Research on Alzheimer's Disease, Columbia University, New York, NY, USA

<sup>3</sup>Institute for Computational Biology, Case Western Reserve University, Cleveland, OH, USA

### **Objectives**

Mutations in *APP*, *PSEN1* and *PSEN2* lead to early-onset Alzheimer disease (EOAD). These mutations account for ~11% of EOAD overall, leaving the majority of genetic risk for the most severe form of Alzheimer disease unexplained.

### **Methods**

To search for rare variants contributing to risk for EOAD we performed Whole-Exome Sequencing (WES) in 50 Caucasian EOAD cases screened negative for *APP*, *PSEN1*, and *PSEN2*. Variant filtering for functional, damaging rare variants (MAF<0.1%) was performed. Damage prediction was performed using a CADD score (Combined Annotation Dependent Depletion; top 10%). Rare, damaging variants shared by multiple cases (+2) were then selected for follow-up protein-protein interaction analysis with known EOAD genes (*APP*, *PSEN1*, *PSEN2*, *GRN*, *MAPT*) using the program STRINGdb. Additionally, we performed a separate text-mining screen of this set of variants using Phenolyzer and the terms 'Alzheimer disease', 'dementia', and 'neurodegeneration'.

### **Results**

We identified 51 rare, damaging variants in 46 genes in two or more EOAD cases. Gene network analysis of these 46 genes with known EOAD genes identified five candidate genes: HSPG2 (interacts with GRN and APP), CLSTN1 (interacts with PSEN1 and APP), and DOCK3 (interacts with PSEN1 and PSEN2). Five cases have a variant in HSPG2, a gene in a LOAD susceptibility region and potentially involved in amyloidogenesis and tau aggregation in AD. Four cases have a variant in DOCK3, a gene shown to regulate amyloid- $\beta$  secretion, and associated with neurofibrillary tangles in AD brains. Several other cases have shared variants in CLSTN1. Disruption of calyntenin-1-associated axonal transport of APP by mutations in CLSTN1, a known APP interactor, have been identified as a potential pathogenic mechanism of Alzheimer's. Moreover, CLSTN1's potential as a regulator of synapse formation and neuronal development suggests other mechanisms through which it could be involved in development of dementia. Top scoring genes for the Phenolyzer analysis included the genes HSPG2, STAT1, a BACE1 interactor, and CNTNAP2, whose expression is down-regulated in AD patients.

### **Conclusions**

WES of EOAD cases identified several genes with potential roles in AD pathogenesis.

## Appendix IX:

### **SORL1 mutations and Parkinsonian features in early onset Alzheimer's disease families**

Cuccaro ML, Carney RM, Kunkle BW, Vance JM, Whitehead PL, Gilbert JR, Vardarajan BN, Haines JL, Mayeux R, Pericak-Vance MA

Early-onset Alzheimer's disease (EOAD; onset of Alzheimer's disease (AD) prior to age 65) affects ~200,000 individuals, representing 3.7% of all AD cases. Mutations in *PSEN1*, *PSEN2*, and *APP* can cause EOAD. We conducted a genome-wide search to identify novel mutations in a series of families enriched for EOAD using whole-exome sequencing (WES). We hypothesized that WES would identify novel mutations in these families and that these mutations would be associated with a spectrum of neurodegenerative features.

Our dataset consisted of 50 families with at least one individual with EOAD. All individuals and additional family members were characterized via comprehensive medical and research record review. WES (Agilent SureSelect) was performed on individuals of interest per family; *APOE* genotypes were available as well.

WES identified variants of interest in *SORL1*, a gene which previously had been primarily associated with late-onset AD. *SORL1* missense mutations were identified in 2 families. One family contained 4 affected individuals (AOO range 59-82 years; 3 with *APOE* genotype 3/4, 1 with 3/3) and 2 unaffected individuals (ages 81-84, both *APOE* 3/3) with the novel T588I *SORL1* mutation. Two individuals with *SORL1* mutations in this family also presented with Parkinsonian features. The second family carried the previously reported *SORL1* mutation (T749M) with 3 affected individuals with the mutation (AOO 55-84, all *APOE* 3/3), 1 affected individual without the mutation (AOO 76, *APOE* 3/4), and 1 unaffected individual with the mutation (age 79, *APOE* 3/3). One affected individual with in the second family with the T749M *SORL1* mutation had neuropathologic evidence of Lewy bodies without clinical Parkinsonism.

A follow-up of *SORL1* mutations in a separate dataset of LOAD families (Vardarajan 2015, *Annals of Neurology*) corroborated the presence of Parkinsonism in 4 individuals with *SORL1* mutations (3 individuals with the common rs2298813 mutation and 1 individual with a rare mutation).

This study confirms *SORL1* as a gene involved in risk for EOAD. Equally compelling is our finding of Parkinson-related features and Lewy bodies without clinical Parkinsonism in association with these *SORL1* mutations. This is the first study to suggest a relationship between *SORL1*, AD, and Parkinsonism.